Def. A pair $(X, \mathcal{R})$, where X is a set of elements and $\mathcal{R}$ is a collection of subsets of X, is called a set system.

(= range space in computational geometry)

infinite set system에서 finite set system으로 가능.

Def. For a set system $(X, \mathcal{R})$ and $Y \subseteq X$, the projection of $\mathcal{R}$ on Y is

$$\mathcal{R}|_Y := \{Y \cap R : R \in \mathcal{R}\}. \qquad (Y, \mathcal{R}|_Y)$$

Example.

(1) Let $X = \mathbb{R}$, $\mathcal{R} = \{\text{intervals}\}$

(2) Let $X = \mathbb{R}^d$, $\mathcal{R} = \{\text{half-planes}\}$.

(3) Let $X = \mathbb{R}^d$, $\mathcal{R} = \{\text{convex polygons}\}$.
$d \geq 2$

Hitting set problem.

Given $(X, \mathcal{R})$, what is the smallest $Y \subseteq X$ that intersects all sets in $\mathcal{R}$?

graph의 경우 transversal

Def.
Vapnik-Chervonenkis dimension (or VC-dim.) of $(X, \mathcal{R})$, denoted by $VC(\mathcal{R})$, is the minimum d s.t.

$|\mathcal{R}|_Y| < 2^{|Y|}$ for any finite $Y \subseteq X$ with $|Y| > d$.

Observation.

VC-dim. is hereditary: VC-dim of $(X, \mathcal{R}) \leq d$

$\Rightarrow \forall Y \subseteq X$, VC-dim of $(Y, \mathcal{R}|_Y) \leq d$. (trivial. $\exists Z \subseteq Y$ s.t. $d' > d$ $\Rightarrow |\mathcal{R}|_{Y}|_{Z}| = 2^{d'}$, ↯.)

e.g. X: Euclidean, Y: finite subset.

Def. $Y \subseteq X$ is shattered by $\mathcal{R}$ if $|\mathcal{R}|_Y| = 2^{|Y|}$.

The shatter function $\pi_{\mathcal{R}}$ of $(X, \mathcal{R})$ is defined by

$$\pi_{\mathcal{R}}(m) := \max\{|\mathcal{R}|_Y| : Y \subseteq X, |Y| = m\}.$$

Lemma (Sauer–Shelah lemma).

$$VC(\mathcal{R}) \leq d \Rightarrow \forall m \geq 1,\ \pi_{\mathcal{R}}(m) \leq \sum_{i=0}^{d} \binom{m}{i} = O(m^d).$$

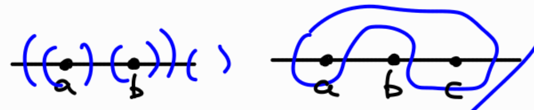
- $m \leq d \Rightarrow \sum_{i=0}^{d} \binom{m}{i} = \sum_{i=0}^{m} \binom{m}{i} = 2^m = \pi_{\mathcal{R}}(m)$.
- $m > d \Rightarrow \pi_{\mathcal{R}}(m) < 2^m$ by the def. of $VC(\mathcal{R}) \leq d$.

  $\pi_{\mathcal{R}}(m) = O(m^d)$ by the lemma.

Example.

(1) Let $X = \mathbb{R}$, $\mathcal{R} = \{\text{intervals}\}$

Then $VC(\mathcal{R}) = 2$, thus $\pi_{\mathcal{R}}(m) = O(m^2)$.

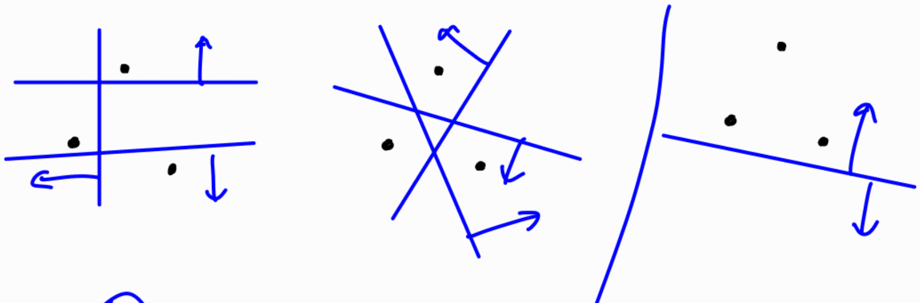
$\because$ 

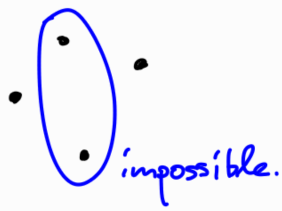
$\{a, b, c\}$ is not shattered.
($\{a,b,c\} \cap \mathcal{R} \not\ni \{a,c\}$.)

$\pi_{\mathcal{R}}(m) = \Theta(m^2)$. (Known. tight)

(2) Let $X = \mathbb{R}^d$, $\mathcal{R} = \{\text{half-planes}\}$.

Then $VC(\mathcal{R}) = d+1$, thus $\pi_{\mathcal{R}}(m) = O(m^{d+1})$.

$\because$ 



$\pi_{\mathcal{R}}(m) = \Theta(m^d)$. (Known) not tight

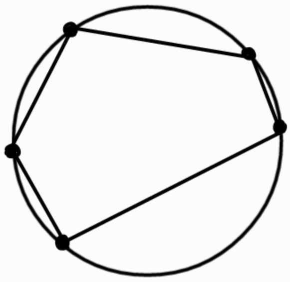
(3) Let $X = \mathbb{R}^d$, $d \geq 2$, $\mathcal{R} = \{\text{convex polygons}\}$.

Then $VC(\mathcal{R}) = \infty$, thus $\pi_{\mathcal{R}}(m) = 2^m$.

$\because$ Let $A \subset \{(x,y) : x^2 + y^2 = 1\}$, $|A| = n$. ($d=2$)

$\forall A' \subseteq A$ with $|A'| = m \leq n$,

$\exists$ convex m-gon G s.t. $A \cap G = A'$.



Def. Given finite $(X, \mathcal{R})$ and $0 \leq \varepsilon \leq 1$, $N \subseteq X$ is an ε-net for $\mathcal{R}$ if $N \cap R \neq \emptyset$ for all $R \in \mathcal{R}$ with $|R| \geq \varepsilon |X|$.

Def. Given a weight func $w : X \to \mathbb{R}^+$ s.t. $w \not\equiv 0$,

$N \subseteq X$ is an ε-net w.r.t. w if $N \cap R \neq \emptyset$ for any $R \in \mathcal{R}$ s.t. $\underbrace{w(R)}_{:= \sum_{x \in R} w(x)} \geq \varepsilon \cdot w(X)$.

e.g. $w(x) = \frac{1}{|X|}$ (X: finite)

$w(x) = 1$

Remark.

weight 있는 version은 일반적인 ε-net 인데 X 의 element들이 multiple copy...

THEOREM 47.4.2 [AS08]

Let $(X, \mathcal{R})$ be a finite set system with $\pi_{\mathcal{R}}(m) = O(m^d)$ for a constant d, and $0 < \epsilon, \gamma \leq 1$ be given parameters. Let $N \subseteq X$ be a set of size

$$\max\left\{\frac{4}{\epsilon}\log\frac{2}{\gamma},\ \frac{8d}{\epsilon}\log\frac{8d}{\epsilon}\right\}$$

chosen uniformly at random. Then N is an ϵ-net with probability at least $1 - \gamma$.

Fact.

$VC(\mathcal{R}) \leq d \Rightarrow \forall \varepsilon > 0$, ε-net of size $O\left(\frac{d}{\varepsilon}\log\frac{d}{\varepsilon}\right)$ can be computed deterministically in poly$(\frac{1}{\varepsilon})|X|$ time.

1 exists

2 deterministically

3 poly$(\frac{1}{\varepsilon})|X|$ time

$f(\frac{1}{\varepsilon})$.

$f(x) = dx \log(dx)$.

THEOREM 47.4.3 [BCM99]

Let $(X, \mathcal{R})$ be a finite set system such that VC-dim$(\mathcal{R}) = d$, and $\epsilon > 0$ a given parameter. Assume that for any $Y \subseteq X$, all sets in $\mathcal{R}|_Y$ can be computed explicitly in time $O(|Y|^{d+1})$. Then an ϵ-net of size $O(\frac{d}{\epsilon}\log\frac{d}{\epsilon})$ can be computed deterministically in time $O(d^{3d}) \cdot (\frac{1}{\epsilon}\log\frac{1}{\epsilon})^d \cdot |X|$.

Example. (sizes of ε-nets)

(1) Let $X = \mathbb{R}$, $\mathcal{R} = \{\text{intervals}\}$ $\quad \leq \frac{1}{\varepsilon}$

(2) Let $X = \mathbb{R}^d$, $\mathcal{R} = \{\text{half-planes}\}$. $\quad \leq \frac{d}{\varepsilon}\log\frac{1}{\varepsilon}$ ($d = 2, 3$: better)

(3) Let $X = \mathbb{R}^d$, $d \geq 2$, $\mathcal{R} = \{\text{convex polygons}\}$.

Objects	SETS	UPPER BOUND		LOWER BOUND	
Intervals	P/D	$\frac{1}{\epsilon}$		$\frac{1}{\epsilon}$	
Lines, $\mathbb{R}^2$	P/D	$\frac{2}{\epsilon}\log\frac{1}{\epsilon}$	[HW87]	$\frac{1}{2\epsilon}\frac{\log^{1/3}\frac{1}{\epsilon}}{\log\log\frac{1}{\epsilon}}$	[BS17]
Half-spaces, $\mathbb{R}^2$	P/D	$\frac{2}{\epsilon} - 1$	[KPW92]	$\frac{2}{\epsilon} - 2$	[KPW92]
Half-spaces, $\mathbb{R}^3$	P/D	$O(\frac{1}{\epsilon})$	[MSW90]	$\Omega(\frac{1}{\epsilon})$	
Half-spaces, $\mathbb{R}^d$, $d \geq 4$	P/D	$\frac{d}{\epsilon}\log\frac{1}{\epsilon}$	[KPW92]	$\frac{\lfloor d/2 \rfloor - 1}{9}\frac{1}{\epsilon}\log\frac{1}{\epsilon}$	[PT13] [KMP16]
Disks, $\mathbb{R}^2$	P	$\frac{13.4}{\epsilon}$	[BGMR16]	$\frac{2}{\epsilon} - 2$	[KPW92]
Balls, $\mathbb{R}^3$	P	$\frac{2}{\epsilon}\log\frac{1}{\epsilon}$		$\Omega(\frac{1}{\epsilon})$	
Balls, $\mathbb{R}^d$, $d \geq 4$	P	$\frac{d+1}{\epsilon}\log\frac{1}{\epsilon}$	[KPW92]	$\frac{\lfloor d/2 \rfloor - 1}{9}\frac{1}{\epsilon}\log\frac{1}{\epsilon}$	[KMP16]
Pseudo-disks, $\mathbb{R}^2$	P/D	$O(\frac{1}{\epsilon})$	[PR08]	$\Omega(\frac{1}{\epsilon})$	
Fat triangles, $\mathbb{R}^2$	D	$O(\frac{1}{\epsilon}\log\log^*\frac{1}{\epsilon})$	[AES10]	$\Omega(\frac{1}{\epsilon})$	
Axis-par. rect., $\mathbb{R}^2$	D	$\frac{5}{\epsilon}\log\frac{1}{\epsilon}$	[HW87]	$\frac{1}{9}\frac{1}{\epsilon}\log\frac{1}{\epsilon}$	[PT13]
Axis-par. rect., $\mathbb{R}^2$	P	$O(\frac{1}{\epsilon}\log\log\frac{1}{\epsilon})$	[AES10]	$\frac{1}{16}\frac{1}{\epsilon}\log\log\frac{1}{\epsilon}$	[PT13]
Union $\kappa_{\mathcal{R}}(\cdot)$, $\mathbb{R}^2$	D	$O(\frac{\log(\epsilon \cdot \kappa_{\mathcal{R}}(1/\epsilon))}{\epsilon})$	[AES10]	$\Omega(\frac{1}{\epsilon})$	
Convex sets, $\mathbb{R}^d$, $d \geq 2$	P	$\lvert X \rvert - \epsilon \lvert X \rvert$		$\lvert X \rvert - \epsilon \lvert X \rvert$	

Black boxes.

$Y \subseteq_{fin} X$

$|Y| - \varepsilon|Y|$

$|Y| = 1억$, $Y|_R = 2^Y$

$\frac{2}{1억}$-net? Y $Y-\{p\}$ ~~$Y-\{p,q\}$~~ does not hit $\{p,q\}$

1. Net finder.

A <u>net finder of size $f(r)$ for $(X, \mathcal{R})$</u> is an algorithm $\boxed{A}$ s.t.

given $r \in \mathbb{R}^+$ and $w : X \to \mathbb{R}_{\geq 0}$,

$\boxed{A}$ returns an $\frac{1}{r}$-net of size $f(r)$ for $(X, \mathcal{R})$ w.r.t. w.

$r \uparrow \Rightarrow$ small-net, $\underset{\text{size}}{f(r)} \uparrow$

2. Verifier.

A <u>verifier</u> is an algorithm $\boxed{B}$ s.t.

given $H \subseteq X$, $\boxed{B}$ (correctly) returns "H is a hitting set",

or $R \in \mathcal{R}$ s.t. $R \cap H = \phi$. (fail하는 witness 알려줌.)

Let $T_{\boxed{A}} = T_{\boxed{A}}(\underbrace{|X|, |\mathcal{R}|}_{\text{depends on only sizes}}, r)$ be the running time of $\boxed{A}$ and

$T_{\boxed{B}} = T_{\boxed{B}}(|X|, |\mathcal{R}|)$ for $\boxed{B}$.

$f(x) = dx\log(dx)$, above Fact.

<u>Fact.</u> If $VC(\mathcal{R}) \leq d$, $T_{\boxed{A}}$ and $T_{\boxed{B}}$ are polynomials.

(<u>constant</u> d is used in the polynomial.)

$\left(\frac{1}{r}\right)^d |X|$ $\qquad$ $|X|^d$ or $|X|^{d+1}$

<u>Theorem.</u>

Let c be the size of the optimal hitting set for $\overset{\text{finite}}{(X, \mathcal{R})}$.

Suppose <u>$T_{\boxed{A}}$ and $T_{\boxed{B}}$ are polynomial.</u> Not very important.

Then $\exists$ algorithm that gives a hitting set of size $\leq f(4c)$

in $O\left(c\log\left(\frac{|X|}{c}\right)\left(T_{\boxed{A}}(|X|, |\mathcal{R}|, c) + T_{\boxed{B}}(|X|, |\mathcal{R}|)\right)\right)$ time.

Algorithm.

Assume that we know the size c of a smallest hitting set.

Strategy : "Survival of the fittest".

Put weights on the elements of X uniformly, 1.

Iterative procedure :

(1) Use $\boxed{A}$ to get a $\frac{1}{2c}$-net $\mathcal{N}$ of size $f(2c)$.

(2) Use $\boxed{B}$ with $\mathcal{N}$.

(3-1) If $\mathcal{N}$ is a hitting set, then done.

If the process iterated k times, then

it took time $k \cdot \left(T_{\boxed{A}}(|X|, |\mathcal{R}|, 2c) + T_{\boxed{B}}(|X|, |\mathcal{R}|) \right)$

$= O\left(k \cdot \left(T_{\boxed{A}} + T_{\boxed{B}}\right)\right)$. ($\because T_{\boxed{A}}$: poly)

(3-2) If $\mathcal{N}$ is not a hitting set,

then $\boxed{B}$ gives $R \in \mathcal{R}$ s.t. $R \cap \mathcal{N} = \emptyset$.

Double the weights of elements in R,

and repeat from (1).

Claim. If there is a hitting set of size c,

above "doubling procedure" cannot iterate

more than $4c \log\left(\frac{n}{c}\right)$ times.

(And $w(X) \leq \frac{n^4}{c^3}$.)

Pf of the claim. (this argument was used in several papers.)

Suppose that k iterations are performed.

In each iteration, if $\boxed{B}$ returns R,

then $w(R) \leq \frac{1}{2c} w(X)$, thus

$w(X)$ is not multiplied by more than $1 + \frac{1}{2c}$.

$\Rightarrow w(X) \leq |X| \cdot \left(1+\frac{1}{2c}\right)^k \leq |X| e^{\frac{k}{2c}}$. $\cdots(*)$ Taylor exp. of e^x

Let H be a hitting set of size c.

Then in each iteration, at least one element of H is doubled. Say each $h \in H$ has been doubled z_h times.

$$\Rightarrow w(H) = \sum_{h \in H} 2^{z_h} \text{ and } \sum_{h \in H} z_h \geq k.$$

Hence $\frac{w(H)}{c} = \mathbb{E}_h[2^{z_h}] \geq 2^{\mathbb{E} z_h} \geq 2^{\frac{k}{c}}$,

Jensen's ineq. $x \mapsto 2^x$: convex.

resulting $w(H) \geq c \cdot 2^{\frac{k}{c}}$. $\cdots(**)$

Since $w(H) \leq w(X)$, by $(*)$ & $(**)$,

$c \cdot 2^{\frac{k}{c}} \leq |X| e^{\frac{k}{2c}} \leq |X| 2^{\frac{3k}{4c}}$. $\left(\because \overset{2.7\sim}{e} \leq 2^{\overset{2.8\sim}{\frac{3}{2}}}\right)$

Thus $k \leq 4c \cdot \log \frac{|X|}{c}$. $w(X) \leq \frac{n^4}{c^3}$ also follows from $(*)$.

$\left(\log c + \frac{4k}{4c} \leq \log|X| + \frac{3k}{4c} \Rightarrow \frac{k}{4c} \leq \log\frac{|X|}{c}\right)$

$$\left(|X| e^{\frac{k}{2c}} \leq |X| e^{2\log\frac{|X|}{c}} \leq |X| 2^{\log\left(\frac{|X|}{c}\right)^3} = |X| \frac{|X|^3}{c^3}\right)$$

$\square$ of Claim.

<u>Last issue</u>: How can we assume that c is known?

Start with a guess, $c' = 1$. (hitting set of singleton.)

If the number of iterations exceeds the bound, then replace c' with $2c'$.

Our conjectured c' will be at most twice the optimal one.

Thus the hitting set obtained is of size $\leq f(2 \cdot 2c)$.

Running-Time.

Given c', the procedure takes

$$4c' \log\left(\frac{|X|}{c'}\right) \cdot \left(T_{\boxed{A}}(|X|, |R|, 2c') + T_{\boxed{B}}(|X|, |R|) \right).$$

polynomial

$\Downarrow$

$$c' : 1 \to 2 \to 4 \to \cdots \to c'$$

$$c \le c' \le 2c$$

last c' dominates running time.

절반 절반

$2c$ worst case

Thus $O\left(c \log\left(\frac{|X|}{c}\right) \left(T_{\boxed{A}}(|X|, |R|, c) + T_{\boxed{B}}(|X|, |R|) \right) \right)$.

Another method : LP + net finder 사실 저번 방식 발표 때 함.

only once. $\boxed{A}$ 느리면 significantly faster.

더 나중 결과. hitting set size $\frac{1}{4}$, 속도 더 느리지 않음.

(10년 뒤)